

IDENTIFICAÇÃO DO ESTADO DE CONDIÇÃO DE TALUDES EM ROCHA ATRAVÉS DA APLICAÇÃO DE FERRAMENTAS ESTATÍSTICAS AVANÇADAS

ROCK SLOPE CONDITION STATE IDENTIFICATION THROUGH THE APPLICATION OF ADVANCED STATISTICAL TOOLS

Tinoco, Joaquim; *ISISE/Universidade do Minho, Guimarães, Portugal, jtinoco@civil.uminho.pt*
Gomes Correia, António; *ISISE/Universidade do Minho, Guimarães, Portugal, agc@civil.uminho.pt*
Cortez, Paulo; *ALGORITMI/Universidade do Minho, Guimarães, Portugal, pcortez@dsi.uminho.pt*
Toll, David; *University of Durham, Durham, UK, d.g.toll@durham.ac.uk*

RESUMO

Neste trabalho é analisada a adequabilidade da utilização de ferramentas de *data mining* (DM) no desenvolvimento de um modelo de classificação do estado de condição de taludes em rocha da rede rodovia e ferroviária. O modelo proposto utiliza informação recolhida durante inspeções de rotina, por norma de fácil obtenção, complementada com informação geométrica, geológica e geográfica do talude e atribui um nível/classe ao estado de condição ao talude em análise. O sistema proposto, desenvolvido através da aplicação de ferramentas de inteligência artificial, procura maximizar a informação disponível visando uma classificação realista do estado de condição do talude. Para a aprendizagem dos modelos foi utilizada uma base de dados de taludes em rocha da rede ferroviária britânica. Os resultados obtidos permitem analisar a capacidade de aprendizagem de algoritmos de DM na análise de problemas de elevada complexidade e inferir sobre a informação necessária para uma correta identificação do real estado de condição de taludes em rocha.

ABSTRACT

In the present work it is analysed the suitability of data mining techniques in the development of a classification model of the condition state of rock slopes of the highway and railway network. The proposed model is fed with information taken from routine inspections, usually easy to obtain, as well as geometric, geologic and geographic information of the slope and labels it with a class according to its condition state. The proposed model, developed based on the application of artificial intelligent tools, intends to maximize all the available information towards a representative classification of the rock slope condition state. For model learning, a data base was used with rock slope data from the UK railway network. The achieved results allow assessing the learning capacity of DM algorithms when applied to the analysis of complex problems, as well as inferring about the information required for a correct identification of the real condition state of a rock slope.

1 - INTRODUÇÃO

Após um longo período de investimento e desenvolvimento, Portugal dispõe atualmente de uma rede de infraestruturas de transporte, nomeadamente rodovia e ferroviária, bastante completa. O desafio atual prende-se com a manutenção da rede existente de forma a assegurar todas as condições de segurança e mobilidade. Face ao elevado número de elementos constituintes da rede e das limitações orçamentais disponíveis para gestão de toda a infraestrutura, torna-se fundamental dispor de um conjunto de ferramentas que auxiliem os gestores responsáveis nas suas tarefas de forma a otimizar os recursos.

Um dos elementos que requerem particular atenção, necessitando de uma observação/manutenção com regularidade é a rede de taludes que constitui a rede rodovia e ferroviária. A falta de manutenção pode levar à ocorrência de deslizamentos/derrocadas com graves perdas económicas e humanas. Por outro lado, a gestão de toda a rede representa um custo significativo para as respetivas concessionárias. Torna-se portanto fundamental desenvolver um conjunto de ferramentas capazes de identificar o estado de condição de um determinado talude preferencialmente através de informação recolhida durante as inspeções de rotina de forma a minimizar custos. Desta forma será possível priorizar intervenções minimizando os custos de manutenção e a ocorrência de acidentes.

Embora alguns sistemas para a previsão da ocorrência de deslizamentos/derrocadas possam ser encontrados na literatura da especialidade, uma parte significativa tem como alvo taludes naturais, não sendo adequados à avaliação de taludes feitos pelo homem. Além disso, estes sistemas apresentam como principal desvantagem o facto de requerem informação por vezes de difícil obtenção, como por exemplo através de ensaios específicos ou equipamentos de monitorização dispendiosos, acrescentando ainda o facto de em alguns casos terem sido desenvolvidos utilizando informação proveniente de casos de estudo muito concretos, limitando um pouco o respetivo domínio de aplicação. Há ainda a sublinhar o facto de

alguns sistemas serem caracterizados por uma forte componente de subjetividade. A título de referência, e em particular para o caso de taludes em rocha, é de sublinhar os sistemas propostos por Cheng e Hoang (2014) onde é apresentado um método de classificação para taludes em rocha que calcula a probabilidade de rotura de um determinado talude. Também Ahangar-Asr *et al.* (2010) propuseram um modelo para a determinação do fator de segurança de taludes em rocha e em solo, através da aplicação de técnicas de *data mining* (DM). Lu e Rosenbaum (2003) e Sakellariou e Ferentinou (2005) fizeram também uso de técnicas de DM mas neste caso para prever diretamente se um determinado talude iria ruir ou não. Todas estas três abordagens, embora tenham conseguido um desempenho satisfatório, apresentam como principal limitação o facto de terem de assumir *a priori* o tipo de rotura do talude. Além disso, foram desenvolvidos utilizando uma base de dados com um número de registos bastante reduzido. Mais recentemente, um novo sistema foi apresentado por Pinheiro *et al.* (2015). O sistema SQI (Slope Quality Index), caracterizado pela sua grande flexibilidade, assenta na avaliação de diferentes fatores que influenciam o comportamento de um talude, sendo aplicável quer a taludes em rocha quer em solo. Através da ponderação dos diferentes fatores obtém-se uma classificação final do talude, representativa do estado de condição do mesmo.

O comportamento de um talude depende de um elevado número de fatores, alguns deles de difícil avaliação (AGC, 2007; Fay *et al.*, 2012). Por outro lado, existem atualmente poderosas ferramentas capazes de explorar grandes volumes de dados e extrair conhecimento útil. Estas ferramentas, usualmente conhecidas por DM, têm sido aplicadas com sucesso em diversas áreas do conhecimento, nomeadamente na área de Engenharia Civil e em particular em geotecnia (Tinoco *et al.*, 2014a; Miranda *et al.*, 2011). Ao nível do estudo da estabilidade de taludes, Gavin e Xue (2009) calcularam o índice de fiabilidade de um determinado talude, bem como a posição do nível freático com recurso a algoritmos genético. Também Wang *et al.* (2005) avaliaram a estabilidade de um talude através da aplicação de redes neuronais artificiais. Duas outras propostas desenvolvidas através da aplicação de máquinas de vetor de suporte foram apresentadas por Cheng *et al.* (2012) e Yao *et al.* (2008).

O presente trabalho, tem como principal objetivo o uso de ferramentas de inteligência artificial no desenvolvimento de um sistema de identificação do estado de condição de taludes em rocha, alimentado por informação recolhida durante inspeções de rotina e complementada com alguma informação geométrica, geológica e geográfica. Para questões de treino/validação do sistema foi utilizada uma base de dados relativa à rede de taludes da rede ferroviária do Reino Unido, disponibilizada pela NetworkRail. O problema em estudo foi abordado de duas formas distintas: como um problema de classificação nominal; convertido num problema de regressão. Em cada uma das abordagens foram aplicados dois algoritmos de DM, nomeadamente as Redes Neuronais Artificiais (RNAs) e as Máquinas de Vetores de Suporte (MVSs).

2 - BASE DE DADOS

Um elemento fundamental para a realização de qualquer estudo assente na aplicação de ferramentas de DM é a existência de uma base de dados representativa do problema em estudo. O modelo proposto para a identificação do estado de condições de taludes em rocha apresentado neste trabalho foi desenvolvido com uma base de dados constituída por 6258 registos, tendo sido disponibilizada pela NetworkRail e é relativa à rede ferroviária de Inglaterra.

O modelo apresentado para a classificação do estado de condição de um determinado talude em rocha, daqui em diante designado por EHC ("Earthwork Hazard Categorization") é alimentado por 114 variáveis. Abaixo são listadas algumas das variáveis consideradas:

- Altura;
- Inclinação;
- Perfil do talude;
- Proteção da superfície;
- Tipo de rocha;
- Comprimento do talude;
- Vegetação das bermas;
- Existência de bermas;
- Existência de árvores;
- Existência de árvores instáveis;
- Existência de arbustos;
- Quantidade de raízes à superfície;
- Existência de descontinuidades;
- Existência de percolação;
- Existência de água à superfície;
- etc.

Adicionalmente às variáveis acima mencionadas e às restantes utilizadas no modelo, muitas outras tem uma forte influência no estudo da estabilidade de taludes, nomeadamente a quantidade de precipitação durante um determinado período de tempo. No entanto, como tal informação não está disponível na base

de dados utilizada no presente estudo, as mesmas não foram consideradas. De sublinhar contudo que um dos objetivos deste estudo passa por tentar desenvolver um sistema de identificação do estado de condição de taludes em rocha utilizando essencialmente informação visual recolhida durante inspeções de rotina e que seja de fácil obtenção.

A cada registo da base de dados está atribuída uma classe EHC, constituída por 4 níveis (A, B, C e DE), onde A representa um bom estado de condição e DE um mau estado de condição. Esta atribuição resulta da experiência dos Engenheiros e Técnicos da NetworkRail e será assumida como representativa do real estado de condição do talude.

A Figura 1 ilustra a distribuição dos 6258 registos da base de dados pelas 4 classes EHC. Da sua análise é possível observar uma distribuição fortemente assimétrica pelas 4 classes a qual terá um efeito significativo no desempenho dos modelos, tal como analisado e discutido mais detalhadamente na secção 4. Embora do ponto de vista de aprendizagem dos modelos tal distribuição assimétrica da informação tenha um efeito negativo, a mesma é representativa da realidade, tendo em conta que é espetável que uma parte significativa dos taludes da rede se encontre num bom estado de condição (classe A), e que apenas alguns apresentem um elevada probabilidade de rotura (classe DE).

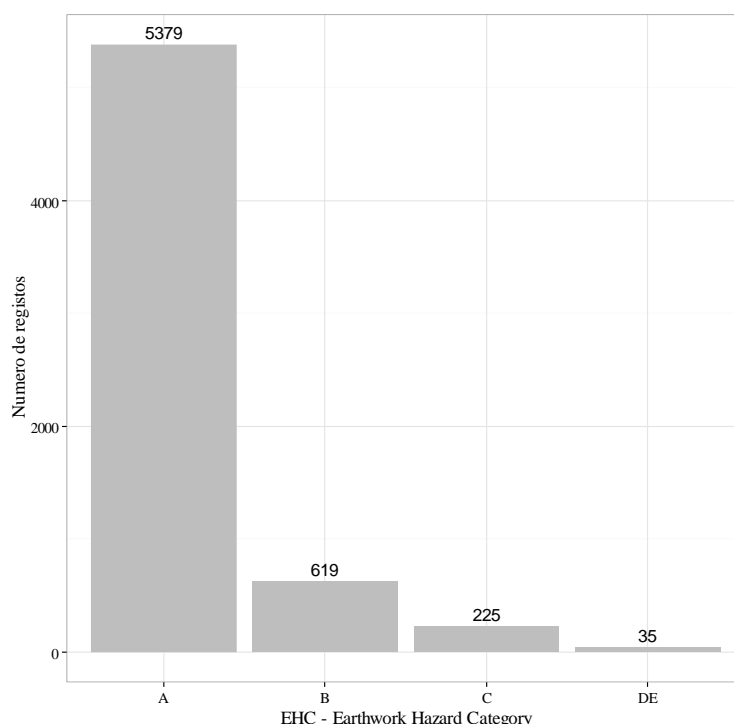


Figura 1 – Distribuição do número de registos pelas 4 classes EHC

3 - METODOLOGIAS E ALGORITMOS

Como referido anteriormente, o sistema proposto para a previsão do EHC foi desenvolvido através da aplicação de ferramentas de DM. Atualmente existem disponíveis diversos algoritmos de DM que podem ser aplicados a problemas de regressão ou de classificação. Neste estudo foram treinados dois algoritmos de referência na abordagem de problemas de regressão e classificação, nomeadamente RNAs e as MVSs. Ambos os algoritmos têm evidenciado elevada eficiência na resolução de problemas reais (Tinoco *et al.*, 2011; Tinoco *et al.*, 2014b; Freitas *et al.*, 2015).

As RNAs procuram imitar alguns aspetos do cérebro humano (Kenig *et al.*, 2001), como o processamento de informação através da interação entre vários neurónios (Perzyk e Kochanski, 2001). Neste trabalho adotou-se o modelo de RNA do tipo unidirecional e multicamada (i.e. *multilayer perceptron*), com uma camada intermédia e com H unidades de processamento. Controlando o valor de H podem ser realizadas análises mais complexas, ainda que um valor elevado de H poderá levar a um sobre-ajustamento do modelo os dados de treino, levando a perda de capacidade de generalização deste. Para ultrapassar esta dificuldade, o valor de H foi definido utilizando uma procura em grelha {0,2,4,6,8}.

As MVS (Cortes e Vapnik, 1995), inicialmente desenvolvidas para problemas de classificação, foram mais tarde também aplicadas a problemas de regressão (Smola e Scholkopf, 2004) após a introdução da função perda ϵ -insensitiva. As MVS apresentam vantagens teóricas sobre as RNA tais como a ausência de mínimos locais durante a fase de aprendizagem. Isto é, estes modelos convergem sempre para a solução ótima. A

ideia subjacente a uma MVS é transformar os dados de entrada num espaço característico de elevada dimensão usando um mapeamento não linear. Posteriormente, a MVS encontra o melhor hiperplano dentro do espaço característico. Esta transformação depende da função kernel adotada. O kernel Gaussiano é o mais popular por apresentar um menor número de parâmetros, tendo sido adaptado no presente estudo. Para auxiliar a escolha dos valores dos diferentes hiperparâmetros (γ , C e ϵ), foram adotadas as heurísticas propostas por Cherkassky e Ma (2004). Assim, para C foi adotado o valor de $C=3$ e a largura da zona ϵ -insensitiva foi definida de acordo com $\epsilon = \hat{\sigma}/\sqrt{N}$, onde $\hat{\sigma} = 1.5/N \times \sum_{i=1}^N (y_i - \hat{y}_i)^2$, \hat{y}_i é o valor previsto pelo algoritmo dos 3-vizinhos próximos e N representa o número de registos da base de dados. O parâmetro kernel γ foi definido usando uma procura em grelha entre $\{1, 3, 5, 7, 9\}$.

Como ilustrado anteriormente na Figura 1, a distribuição dos registos pelas 4 classes EHC é fortemente assimétrica, aspeto que tem um efeito preponderante no desempenho dos algoritmos de aprendizagem. No sentido de ultrapassar esta questão, foram aplicadas duas abordagens visando o balanceamento da base de dados antes de iniciar o processo de aprendizagem. Assim, foram aplicadas as abordagens SMOTE (Synthetic Minority Over-sampling Technique) e Oversampling. O SMOTE (Chawla *et al.*, 2002) permite criar uma “nova” base de dados através da criação de novos registos tendo por base registos semelhantes (k vizinhos próximos). Esta estratégia é aplicada à classe minoritária. Simultaneamente são também removidos alguns registos da classe maioritária. Embora o SMOTE seja uma abordagem direcionada a problemas de classificação, Torgo *et al.* (2015) adaptaram esta metodologia a problemas de regressão. O Oversampling corresponde a uma simplificação do SMOTE onde, aleatoriamente, registos da classe minoritária são repetidos para que todas as classes fiquem com o mesmo número de registos.

A avaliação do desempenho dos modelos, foi realizada através do cálculo de diferentes métricas, nomeadamente (Baía, 2015): erro de classificação (EC); pontuação média de utilidade (PMU); precisão e exatidão. A PMU permite bonificar ou penalizar uma determinada classe em detrimento de outra. Assim, para o cálculo da PMU considerou-se a seguinte matriz custo-benefício:

Quadro 1 – Matriz custo-benefício				
Observado/Previsto	A	B	C	DE
A	1	-4	-8	-16
B	-2	1	-4	-8
C	-4	-2	1	-4
DE	-8	-4	-2	1

A ideia subjacente à matriz apresentada no Quadro 1 consiste em penalizar qualquer classificação não correta, distinguindo se o erro é por excesso ou defeito. Por exemplo, prever um registo como DE quando ele é A (penalização de -8) é menos penalizado em comparação a uma previsão de A quando o real estado é DE (penalização de -16). À exceção do EC quanto maior o valor da métrica melhor o desempenho do modelo. A capacidade de generalização dos modelos foi avaliada através da aplicação de uma validação cruzada com 5-fold (Hastie *et al.*, 2009) e repetição de cada experiência 20 vezes.

Todas as experiências foram conduzidas no ambiente estatístico R (R Development Core Team, 2009), com o auxílio da biblioteca rminer (Cortez, 2010), a qual é particularmente adequada para o treino dos algoritmos RNAs e MVSs.

4 - RESULTADOS

4.1 - Classificação nominal

Abordando a previsão do EHC como um problema de classificação, os algoritmos RNAs e MVSs foram treinados com a base de dados original mas também com uma base de dados balanceada (igual número de registos por classe) resultante da aplicação das abordagens SMOTE e Oversampling.

O Quadro 2 compara o desempenho dos algoritmos RNAs e MVSs na previsão de EHC através de diferentes métricas. É também comparada a influência do balanceamento da base de dados através das abordagens SMOTE e Oversampling.

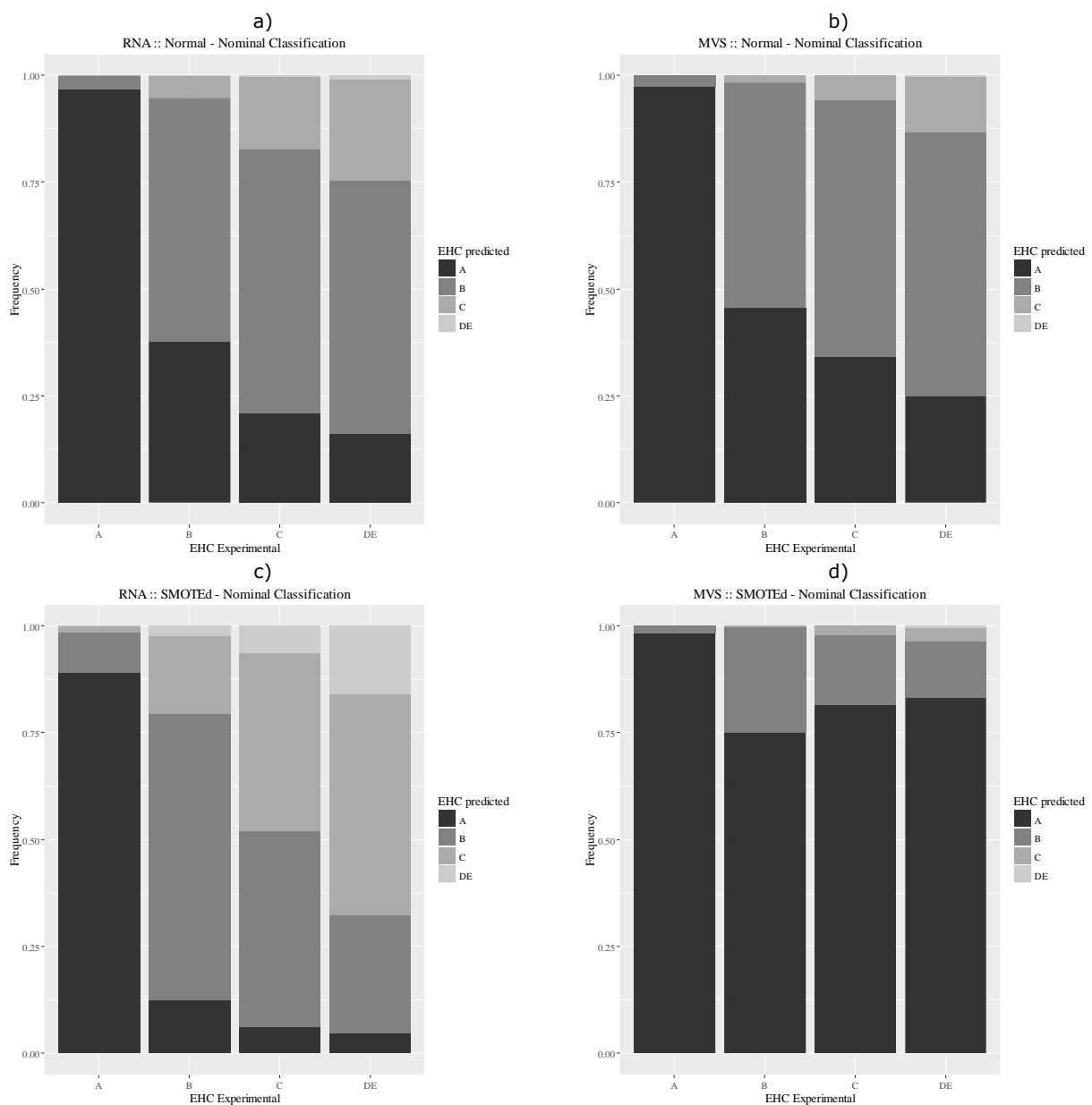
Quadro 2 – Métricas de desempenho dos modelos – classificação nominal

Modelo		EC	PMU	Exatidão & Precisão			
				A	B	C	DE
RNA	Normal	10.56	0.48	96.79 & 94.79	56.94 & 52.16	17.02 & 44.18	0.86 & 9.09
	SMOTed	15.27	0.42	89.05 & 98.08	66.77 & 39.81	41.53 & 32.24	16.00 & 12.16
	OVERed	15.18	0.42	89.09 & 98.01	66.74 & 40.00	43.20 & 33.87	16.29 & 11.62
MVS	Normal	10.95	0.42	97.30 & 93.43	52.72 & 52.10	5.58 & 42.98	0.43 & 30.00
	SMOTed	12.96	0.21	98.35 & 88.65	24.45 & 53.94	2.27 & 56.98	0.57 & 28.57
	OVERed	13.53	0.12	99.88 & 86.54	6.24 & 77.74	0.00 & NA	0.00 & 00.00

A Figura 2 compara o desempenho dos modelos RNAs e MAVs, bem como o efeito do balanceamento da base de dados através das abordagens SMOTE e Oversampling. Em cada um dos gráficos, as barras representam as classes observadas e o preenchimento corresponde à classe prevista pelo modelo. Por exemplo, na Figura 2a, pouco menos de 25% dos registos da classe C foram corretamente previstos como pertencendo a classe C, mais de 50% foram classificados com a classe B e cerca 25% como pertencendo à classe A. É ainda possível observar que uma pequeníssima percentagem foi classificada como DE.

Da análise do Quadro 2 e da Figura 2 observa-se, por um lado que ambos os algoritmos apresentam visíveis dificuldades em identificar corretamente o estado de condição de taludes em rocha. Embora se observe um excelente desempenho na classificação de taludes da classe A, tendo-se conseguido um performance bem perto dos 100% (ver Figura 2a e Figura 2b), para as classes minoritárias, em particular a classe DE, o desempenho dos modelos é muito reduzido. Apenas cerca de 13% dos registos classificados como DE foram corretamente previstos de acordo com o modelo RNA e após balanceamento da base de dados.

Analisando o efeito das abordagens SMOTE e Oversampling, observa-se um desempenho ligeiramente superior na identificação de talude nas classes C e DE após o balanceamento da base de dados, em particular com a utilização das RNAs. Embora positivo, este aumento de rendimento para as classes minoritárias continua a não ser suficiente, tendo inclusive como contrapartida uma pequena diminuição na capacidade de previsão das classes A e B.



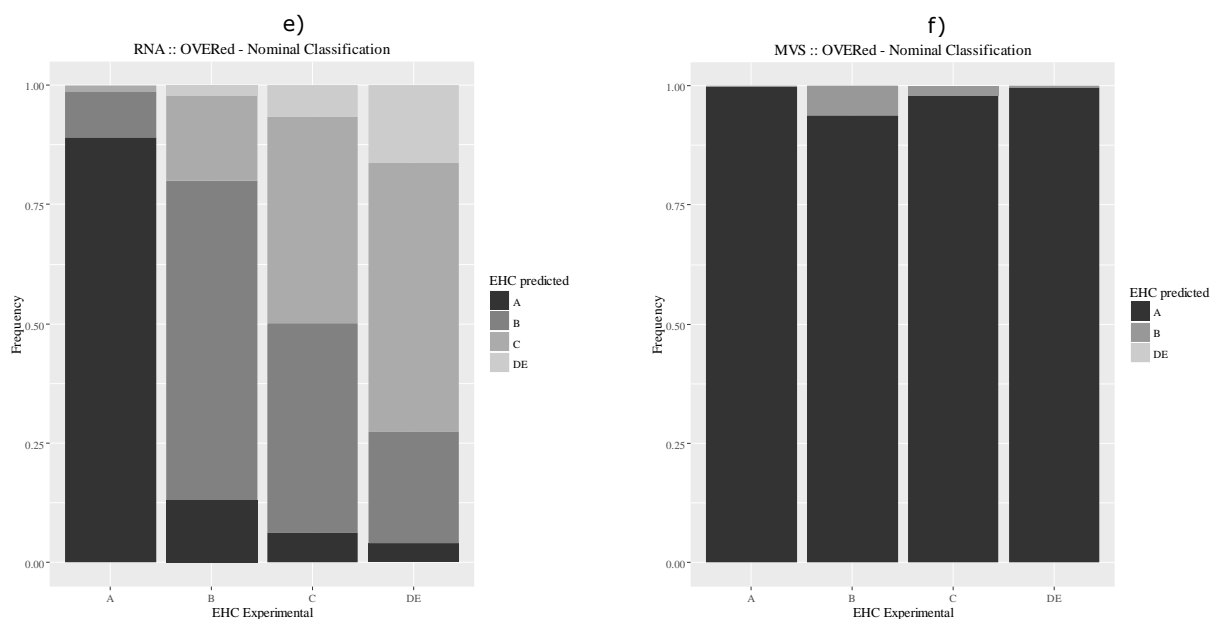


Figura 2 – Comparação do desempenho dos modelos – classificação nominal

4.2 - Regressão

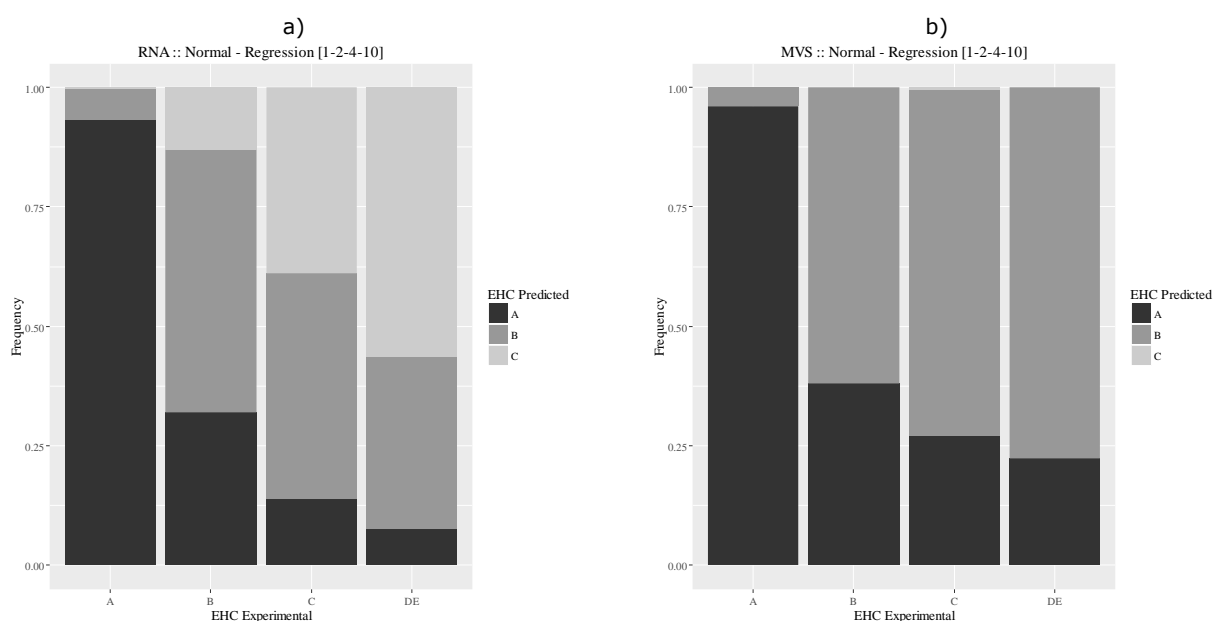
Tendo como objetivo aumentar o desempenho dos modelos, o problema em estudo foi convertido num problema de regressão e resolvido como tal. Para o efeito foram selecionadas diferentes escalas de regressão, tendo-se no final adotado a escala A=1, B=2, C=4 e DE=10.

O Quadro 3 e a Figura 3 mostram e comparam o desempenho dos modelos de regressão na previsão do EHC de taludes em rocha. Uma vez mais observa-se que uma grande dificuldade dos modelos em identificar corretamente o estado de condição do talude, principalmente para as classes correspondentes a estados de condição com uma maior probabilidade de rotura (classes C e DE).

Também aqui o balanceamento da base de dados através da aplicação da abordagem SMOTE para regressão não contribui para uma melhoria na capacidade de previsão das classes minoritárias.

Quadro 3 – Métricas de desempenho dos modelos – regressão

Modelo		EC	PMU	Exatidão e Precisão			
				A	B	C	DE
RNA	Normal	13.04	0.43	93.21 & 95.59	54.96 & 42.43	38.96 & 41.53	0.00 & NA
	SMOTed	18.18	0.44	85.51 & 98.02	66.31 & 32.99	48.78 & 34.20	0.57 & 100
MVS	Normal	11.23	0.45	96.11 & 94.45	62.00 & 49.03	0.64 & 93.55	0.00 & NA
	SMOTed	14.92	0.35	88.23 & 98.66	87.08 & 39.17	17.42 & 54.79	0.00 & NA



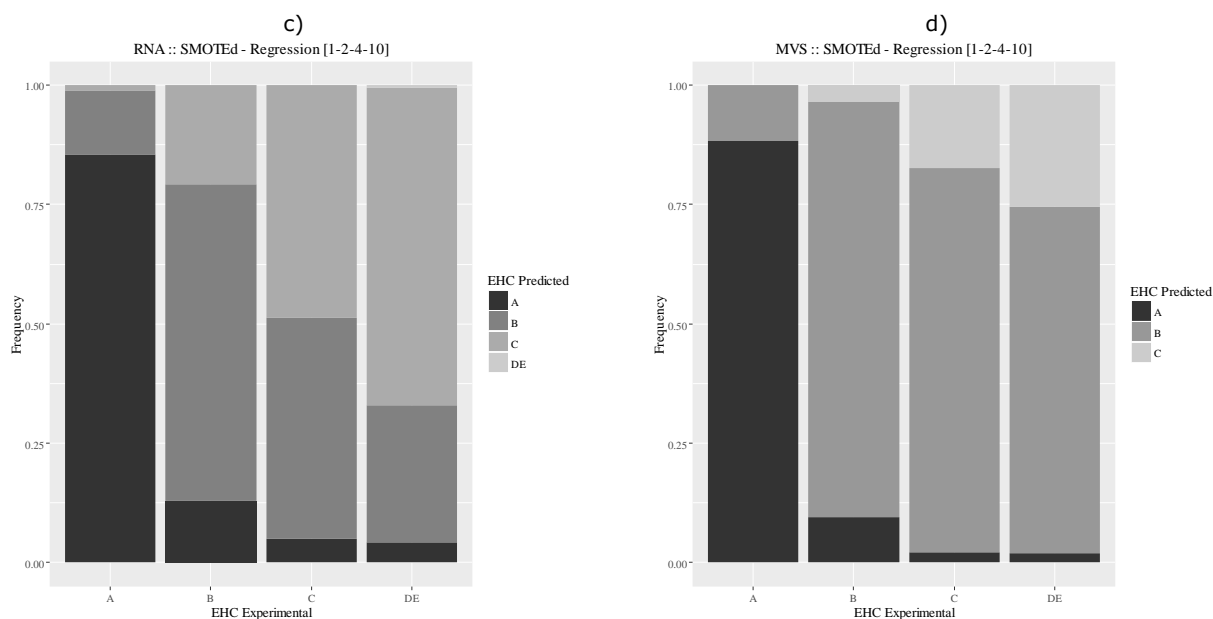


Figura 3 – Comparação do desempenho dos modelos – regressão

5 - OBSERVAÇÕES FINAIS

No presente trabalho foi apresentada uma proposta de identificação do estado de condição de taludes em rocha através da aplicação de ferramentas de análise estatística avançada, usualmente conhecidas por *data mining* (DM). Em particular, foram utilizados dois algoritmos caracterizados pela sua elevada flexibilidade de aprendizagem e com excelentes resultados na resolução de diversos problemas em diferentes áreas do conhecimento, nomeadamente as Redes Neurais Artificiais (RNAs) e as Máquina de Vetores de Suporte (MVSs).

Apesar das diferentes tentativas/abordagens realizadas, os modelos apresentados estão ainda longe de conseguir identificar corretamente o real estado de condição de taludes em rocha. Uma forte dificuldade foi manifestada, quer pelos modelos de classificação nominal, quer pelos modelos de regressão, em prever corretamente o estado de condição de taludes em rocha em particular quando estes se encontram num estado de condição correspondente a uma maior probabilidade de rotura.

Se em parte a distribuição fortemente assimétrica dos registos pelas 4 classes consideradas na avaliação do estado de condição de taludes em rocha poderá estar a condicionar o desempenho dos modelos, é também de equacionar a existência de outros fatores. Por um lado, a ausência de determinada informação, como por exemplo as condições climáticas e hidrogeológicas, poderá estar a condicionar os modelos de aprender corretamente o problema em estudo. Por outro lado, é também de realçar que neste estudo foi aplicada uma aprendizagem supervisionada, onde o real estado de condição dos taludes foi definido por uma equipa de Engenheiros e Técnicos especialistas. Esta situação leva a considerar que a classificação atribuída a cada talude poderá não ser representativa do real estado do talude.

Face a todas estas condicionantes, é de equacionar a aplicação de uma aprendizagem não supervisionada bem como a exploração de outras ferramentas/algoritmos. Será também importante perceber se a classificação atribuída pelos Engenheiros e Técnicos especialistas é representativa do real estado de condição do talude através da confrontação da classificação atribuída nos casos em que ocorreu rotura do talude.

AGRADECIMENTOS

Este trabalho foi financiado pela FCT - "Fundação para a Ciência e a Tecnologia", no âmbito do ISISE, projeto: UID/ECI/04029/2013 e no âmbito do projeto: UID/CEC/00319/2013, bem como através da bolsa de pós-doutoramento com a referência SFRH/BPD/94792/2013 (POCH e FSE). Este trabalho foi também financiado pelo COMPETE: POCI-01-0145-FEDER-007043.

REFERÊNCIAS

- AGC (2007). A national landslide risk management framework for Australia. *Australian Geomechanics Society*, Vol. 42, No. 1, pp. 1-12.
- Ahangar-Asr, A., Faramarzi, A. e Javadi, A. (2010). A new approach for prediction of the stability of soil and rock slopes. *Engineering Computations*, Vol. 27, No.7, pp. 878-893.
- Baía, L. (2015). Actionable forecasting and activity monitoring: applications to financial trading. *Master thesis*. Faculdade de Ciências, Universidade do Porto, Porto, Portugal.
- Chawla, N., Bowyer, K., Hall, L. e Kegelmeyer, W. (2002). Smote: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, pp. 321-357.
- Cheng, M. e Hoang, N. (2014). Slope collapse prediction using Bayesian framework with k-nearest neighbor density estimation: Case study in Taiwan. *Journal of Computing in Civil Engineering*, pp. 1-8.
- Cheng, M., Roy, A. e Chen, K. (2012). Evolutionary risk preference inference model using fuzzy support vector machine for road slope collapse prediction. *Expert Systems with Applications*, Vol. 39, No.2, pp. 1737-1746.
- Cherkassky, V. e Ma, Y. (2004). Practical Selection of SVM Parameters and Noise Estimation for SVM Regression. *Neural Networks*, Vol. 17, No.1, pp. 113-126.
- Cortes, C., e Vapnik, V. (1995). Support Vector Networks. *Machine Learning*, Vol. 20, No.3, pp. 273-297.
- Cortez, P. (2010). Data Mining with Neural Networks and Support Vector Machines using the R/rminer Tool. In P. Perner (Ed.), *Advances in Data Mining - Applications and Theoretical Aspects, 10th Industrial Conference on Data Mining*, LNAI 6171, Springer, pp. 572-583, Berlin, Germany.
- Fay, L., Akin, M. e Shi, X. (2012). Cost-effective and Sustainable Road Slope Stabilization and Erosion Control. *Transportation Research Board*, vol. 430.
- Freitas, E., Tinoco, J., Soares, F., Costa, J., Cortez, P., e Perreira, P. (2015). Modelling tyre-road noise with data mining techniques. *Archives of Acoustics*, Vol. 40, No.4, pp. 547-560.
- Gavin, K. e Xue, J. (2009). Use of a genetic algorithm to perform reliability analysis of unsaturated soil slopes. *Géotechnique*, Vol. 59, No.6, pp. 545-549.
- Hastie, T., Tibshirani, R. e Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, 2nd ed. Springer-Verlag, 745 p.
- Kenig, S., Ben-David, A., Orner, M. e Sadeh, A. (2001). Control of properties in injection molding by neural networks. *Engineering Applications of Artificial Intelligence*, Vol. 14, pp. 819-823.
- Lu, P. e Rosenbaum, M. (2003). Artificial neural networks and grey systems for the prediction of slope stability. *Natural Hazards*, Vol. 30, No.3, pp. 383-398.
- Miranda, T., Correia, A.G., Santos, M., Sousa, L.R. e Cortez, P. (2011). New models for strength and deformability parameter calculation in rock masses using data-mining techniques. *International Journal of Geomechanics*, Vol. 11, pp. 44-58.
- Perzyk, M. e Kochanski, A. (2001). Prediction of ductile cast iron quality by artificial neural Networks. *Journal of Materials Processing Technology*. Vol. 109, pp. 305-307.
- Pinheiro, M., Sanches, S., Miranda, T., Neves, A., Tinoco, J., Ferreira, A. e Gomes Correia, A. (2015). A new empirical system for rock slope stability analysis in exploitation stage. *International Journal of Rock Mechanics and Mining Sciences*, Vol. 76, pp. 182-191.
- R DEVELOPMENT CORE TEAM (2009). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. Web site: <http://www.r-project.org/>, acedido em 28/02/2016.
- Sakellariou, M. e Ferentinou, M. (2005). A study of slope stability prediction using neural networks. *Geotechnical & Geological Engineering*, Vol. 23, No.4, pp. 419-445.
- Smola, A. e Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, Vol. 14, pp. 199-222.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2011). Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time. *Construction and Building Materials*, Vol. 25, No.3, pp. 1257-1262.

- Tinoco, J., Gomes Correia, A. e Cortez, P. (2014a). A novel approach to predicting Young's modulus of jet grouting laboratory formulations over time using data mining techniques. *Engineering Geology*, Vol. 169, pp. 50-60.
- Tinoco, J., Gomes Correia, A. e Cortez, P. (2014b). Support vector machines applied to uniaxial compressive strength prediction of jet grouting columns. *Computers and Geotechnics*, Vol. 55, pp. 132-140.
- Torgo, L., Branco, P., Ribeiro, R. e Pfahringer, B. (2015). Resampling strategies for regression. *Expert Systems*, Vol. 32, No.3, pp. 465-476.
- Wang, H., Xu, W. e Xu, R. (2005). Slope stability evaluation using back propagation neural networks. *Engineering Geology*, Vol. 80, No.3, pp. 302-315.
- Yao,X., Tham, L. e Dai, F. (2008). Landslide susceptibility mapping based on support vector machine: a case study on natural slopes of Hong Kong, China. *Geomorphology*, Vol. 101, No.4, pp. 572-582.